

Cahier des charges du Projet Netflix

- **Contexte : ce que le client nous demande**

Société Netflix gère leurs catalogues de films et séries au travers d'un fichier excel et manuellement et souhaitent moderniser son système d'informations pour accéder plus facilement à leurs données avec une interface graphique.

- **Objet :**

Ce que l'on fait :

- Préviation des scénarios de recherche
- Modélisation d'une base de donnée
- Création d'une base de donnée relationnelle
- Nettoyage des données
- Insertion des données
- Datavisualisation via dashboard

Ce que l'on ne fait pas :

- CRUD
- Détermination des droits d'administration
- Programmation d'une mise à jour

- **Domaine métier :**

Règles de gestion :

- Un film/série peut avoir un ou plusieurs directeurs
- Un film/série peut appartenir à plusieurs catégories
- Un film/série peut être réalisé dans un ou plusieurs pays
- Chaque film/série est identifié par un identifiant unique
- Un film/série peut avoir un ou plusieurs acteurs
- Un acteur peut jouer dans un ou plusieurs films
- Un directeur peut réaliser un ou plusieurs film

- **Critères d'achèvement**

Etape 1 :

On considère l'étape achevée une fois que les données propres, seront interrogeables dans la base de données via des requêtes SQL, ce de manière exhaustive, reproductible et pertinente vis-à-vis des critères de recherche.

Etape 2 :

On considère l'étape achevée une fois que le graphique pertinent avec les données soit visible et compréhensible pour l'utilisateur.

- **Fonctionnement**

Les analyses et traitement des données effectués avec Python et sa librairie pandas.

Conception d'une base de données relationnelle pour le SGBD MySQL en appliquant la méthode MERISE.

Intégration des données dans la base de données avec Python et la librairie SQLAlchemy.

Data Visualisation avec Python et les librairies Matplotlib et Seaborn en les intégrant dans une solution WEB type page HTML avec la librairie Jinja2.

- **Méthodologie :**

Afin d'assurer une intégrité et une structuration cohérente de la base de données, nous avons opté pour la méthodologie MERISE.

Quant à la gestion de projet, nous avons décidé d'adopter un modèle de travail collaboratif type "Waterfall", qui nous permet de séquentialiser un schéma commun de réflexion et de travail. Étant donné la marge de liberté laissée par le commanditaire, cette méthodologie nous a paru opportune de par sa facilité de mise en place et sa rigueur.

- **Planning :**

Date de démarrage : 07/09/2020

Date de fin : 21/09/2020

Recensement des unités d'oeuvre :

Tâche	Dépendances des tâches	Durée jours	Objectifs	Critères d'achèvement	Livrables
Recensement des besoins	-	1.5	Regrouper et énumérer l'ensemble des données	Le besoin client a été formalisé et formulé	Cahier des charges
Dictionnaire des données	1	0.5	Répertorier les données ainsi que leurs définitions	Toutes les données sont recensées	Base de données brutes SQL
Modèle logique de données	-	0.5	Modélisation de la structure de la base de données: définition des entités, et de leurs relations	Tous les attributs sont classé dans une tables avec clés primaire - étrangère	Tables d'entités
Modèle relationnel de données	3	0.5	Modéliser les données	Possibilité d'implémenter des données	Schéma logique
Modèle conceptuel de données	4	1	Mettre en relation les données: formes normales, et établissement des clés primaires et secondaires / Finaliser la base de données	Les modalités relationnelles entre tables et les tableaux relationnelles ont été établies	Schéma relationnel
Modélisation physique sur SQL	5	1	Créer des dataframes propres correspondants en tables de la base de données	Les tables principales sont consultables via conceptuelles / SDSD	Schéma conceptuel / RERE
Création et mise en forme des dataframes pour l'insertion des données	7	0.5	Nettoyer les données	Les données sont standardisées	Table SQL
Data cleaning	-	1	Choix des indicateur statistique défini	Les commandes d'exécution statistique et graphique sont définies	Tronc principal de la base de données SQL
Prospection statistique	7	1	Etablir les relations entre tables	Les tableaux relationnelles sont consultables via Workbench	Base de données cohérente et exhaustive

Arrangement de la base de données sur Workbench, via python	7	2	Tests unitaires sur la base de de données SQL	Corriger les erreurs, éditer les colonnes	Indicateurs statistiques, graphiques exploitables et reproductibles
CRUD	9	3	Intégration du code Python dans le code HTML	Le code HTML est intégré avec le code CSS. L'interface est implémenté	Tableaux relationnelles de la BD
Création de l'interface HTML, via la bibliothèque Jinja	-	5	Représenter avec un tableau de bord, les résultats obtenus	La vue générale est garantie	Tests unitaires
Data visualisation	*	2	Évaluation de la validité du projet	Conformité entre l'application et les attentes du client	Interface utilisateur html
Test utilisateur	*	2	Création d'un deuxième template avec un formulaire	DF	Version finale du produit

Planning prévisionnel:

[illegible]